

Serious Gaming – Spiele als experimentgestützte Evaluationsmethode

Fabian Lang
Tim Püschel
Dirk Neumann

Veröffentlicht in:
Multikonferenz Wirtschaftsinformatik 2012
Tagungsband der MKWI 2012
Hrsg.: Dirk Christian Mattfeld; Susanne Robra-Bissantz



Braunschweig: Institut für Wirtschaftsinformatik, 2012

Serious Gaming – Spiele als experimentgestützte Evaluationsmethode

Fabian Lang

Helmut-Schmidt-Universität, 22043 Hamburg, E-Mail: fabian.lang@hsu-hamburg.de

Tim Püschel

Albert-Ludwigs-Universität, 79098 Freiburg i. Br., E-Mail: tim.pueschel@is.uni-freiburg.de

Dirk Neumann

Albert-Ludwigs-Universität, 79098 Freiburg i. Br., E-Mail: dirk.neumann@is.uni-freiburg.de

Abstract

Diese Arbeit präsentiert das Konzept des Serious Gaming, eine experimentelle Evaluationsmethode. Serious Gaming kann traditionelle Evaluationsmethoden ergänzen und das Evaluationsergebnis verbessern. Web 2.0 bietet hierbei großes Potenzial zur Probandenakquirierung. In dieser Arbeit wird das Konzept erläutert, die Ergänzungsmöglichkeiten traditioneller Methoden aufgezeigt, ein Leitfaden für die Konstruktion eines Serious Game entwickelt, die Distribution im Web 2.0 erläutert und das Konzept anhand einer Fallstudie illustriert. Die Fallstudie offenbart empirische Ergebnisse zur Eignung des Web 2.0 zur Spieldistribution.

1 Einleitung

Es existieren vielerlei Parallelen zwischen realen Prozessen und dem Spiel. So wird beides durch zwischenmenschliche Interaktion und exogene Ereignisse bestimmt. Das Spiel versucht dabei in erster Linie eine authentische Situation widerzuspiegeln.

Das Ziel der Konstruktionsforschung ist die Entwicklung von Artefakten sowie deren Evaluation, die nicht nur technische Faktoren, sondern auch Verhalten untersuchen muss [5]. Spiele können zur Simulation von Umwelten eingesetzt werden und so Verhalten von Menschen untersuchen. In diesem Sinne wurden Kriegsspiele von Offizieren schon seit Jahrtausenden zur Darstellung von Schlachten und Verbesserung ihrer strategischen Fähigkeiten eingesetzt wie bspw. im indischen Brettspiel Chaturanga, das ca. 2000 v.Ch. entwickelt wurde [19]. Im Zweiten Golfkrieg benutzte das Pentagon ein kommerzielles Kriegsbrettspiel, um Szenarien und potenzielle Reaktionsstrategien zu untersuchen [3]. Das Spiel als soziale Interaktion, früher auf örtliche Mehrspielerspiele beschränkt, hat sich in den letzten Jahren stark weiterentwickelt. Anwendungen des Web 2.0 wie Soziale Netzwerke ermöglichen soziale Interaktion und Wettbewerb bei Spielen ohne direkten Kontakt der Spieler. Crowdsourcing ermöglicht es, dass eine große Menge an Spielern einfach akquiriert werden kann.

Die Forschungsfrage adressiert die Anwendung des Konzeptes des Serious Gaming auf Artefakt-Evaluationen, die Konzeption und Implementierung von Serious-Gaming-Experimenten sowie deren Distribution mittels Web 2.0. Im Anschluss an diese Einleitung wird die Evaluationsmethode präsentiert und diskutiert sowie ein Leitfaden zur Konzeption entwickelt. Daraufhin wird die Distribution mittels Web 2.0 erörtert und eine Fallstudie vorgestellt, welche die Anwendung des Leitfadens illustriert sowie zur empirischen Untersuchung der Distribution mittels Web 2.0 dient. Abschließend wird ein Fazit gezogen.

2 Serious Gaming als Forschungsmethode

2.1 Definition

Der Begriff des Serious Gaming (engl. „ernsthaftes Spielen“) wirkt bei oberflächlicher Betrachtung als Paradoxon, da Spielen in der Regel als ein Gegenteil der Ernsthaftigkeit gesehen wird. Es lassen sich jedoch viele Parallelen zwischen der wirklichen Welt und dem Spiel ziehen. Die Akteure erwarten sowohl bei realen Aktionen als auch beim Spiel einen Nutzen aus ihren Taten, sei es durch Belohnung, wie monetäre Auszahlungen, oder durch den symbolischen Sieg und die damit verbundene Zustimmung oder Ehre [6]. Serious Games verbinden das Ernste mit dem Spiel, indem die Spiele einen ernsthaften Zweck verfolgen und dadurch das Spielerische mit der Realität verbinden [16]. Diese Ideen führen zu folgender Definition eines Serious Game (basierend auf [21]):

Ein Serious Game ist ein mentaler Wettbewerb, welcher in Einverständnis mit konkreten Regeln gespielt wird, um ein bestimmtes Ergebnis zu erzielen oder Verhalten der Spieler zu beobachten. Die Absicht eines Serious Game besteht dabei nicht primär in der unterhaltenden Wirkung, sondern vielmehr darin, einen höheren Zweck zu erfüllen.

Anwendung in der Forschung haben Serious Games bisher in der Managementforschung gefunden. Planspiele stellen bestimmte Geschäfts- und Marktmodelle dar, in denen der Spieler als Beobachtungsobjekt Entscheidungssituationen absolvieren muss. Der Vorteil der Planspiele ist, dass sequentielle, mehrdimensionale Probleme untersucht werden können, ohne dass die Ergebnisse zu normativen Aussagen bzw. Präskriptionen ohne Vorhersagekraft für tatsächliches Managementverhalten mutieren [1].

2.2 Einsatzmöglichkeiten in der Konstruktionsforschung

Serious Games bieten sich als Methode in der Konstruktionsforschung an. Der höhere Zweck dieser Spiele kann sich in drei Zielen äußern:

Schaffung eines Referenzpunktes

Bei neuartigen Artefakten ist es möglich, dass keine validen Vergleichsmöglichkeiten existieren. Bspw. kann bei der Entwicklung von Entscheidungsheuristiken vergleichbare Methoden oder Wissen über das Verhalten von Agenten fehlen, sodass keine Simulationen, analytische Vergleichsmethoden oder Szenarien für die Evaluation verwendet werden können. Hier kann ein Serious Game einen Referenzpunkt schaffen, da der Spieler vor eine vergleichbare Entscheidungssituation wie die Heuristik gestellt wird. Die Entscheidungen des Spielers werden durch Logik unter Anwendung von *menschlicher Heuristiken* getroffen. Die menschliche Logik und Intelligenz repräsentiert Vernunft, nicht Rationalität, und kann als Benchmark für Heuristiken verwendet werden.

Aneignung von erfolgreichen Heuristiken

Erfolgreiche Strategien der Spieler können beobachtet und ausgewertet werden, um diese zur Verbesserung des Artefaktes zu nutzen. Hier besteht eine Verwandtschaft zu Planspielen, da diese ebenfalls auf Beobachtung von Entscheidungen abzielen. Durch eine große Anzahl an Spielern kann eine Vielzahl von Strategien und Heuristiken „erspielt“ werden, zudem besteht eine erweiterte kreative Basis. Serious Gaming kann einen dezentralen Suchprozess nach guten Lösungen darstellen (vgl. Weisheit der Vielen: [18]).

Erklärung bzw. Prognose von menschlichem Handeln

Eine Evaluation durch ein Serious Game kann auch Phänomene prognostizieren, die nicht durch konventionelle Methoden vorhersehbar sind. Gerade die gelockerte Rationalitätsannahme ist schwierig zu modellieren und zu prognostizieren. Da das Evaluationsergebnis durch Entscheidungen von echten Menschen bestimmt wird, prognostizieren sie gleichzeitig reales Verhalten von Menschen sehr zutreffend. So könnte man zwar z.B. das Verhalten eines rationalen Individuums bei einer Auktion spieltheoretisch beschreiben, allerdings keine Rückschlüsse auf das Verhalten der echten Auktionsteilnehmer ziehen. Ein Auktionsmechanismus könnte mit dieser Methodik nicht ohne weitere Prüfung, wie eine Feldstudie, bewertet werden. Der verhaltenswissenschaftliche Ansatz eignet sich gut für die Untersuchung der Interaktion zwischen Mensch und Informationssystem bzw. zwischen verschiedenen Menschen. Dies ist ein notwendiger Prozess in der Konstruktionsforschung, da Technologie und Verhalten nicht getrennt werden können und gründliche Forschung beidem bedarf [5].

3 Ergänzung traditioneller Evaluationsmethoden

Serious Gaming bietet vielerlei Anknüpfungspunkte an traditionelle Evaluationstechniken. Üblicherweise werden bei Evaluationen Simulationen, analytische Ansätze, Feldstudien oder Laborexperimente eingesetzt [11]. Jedoch weisen diese Methoden auch Schwachstellen auf, die durch Serious Gaming komplementiert werden können. Zudem sind alternative Ansätze essentiell für eine gründliche Evaluation [5].

Theoretische Modelle bzw. spieltheoretische Ansätze sind eine abstrakte Näherung, um eine ökonomische Intuition des Problems zu erlangen und entscheidende von unwichtigen Faktoren zu separieren. Der Kern theoretischer Modelle ist die formelle Repräsentation, die mit starken Annahmen, Beschränkungen und Verallgemeinerungen verknüpft ist. Es wird zwischen zwei Arten von Vereinfachungen unterschieden: Fiktion und grenztypische Vereinfachungen. Insbesondere die grenztypischen Vereinfachungen, wie vollständige Informationen oder vollkommene Konkurrenz, sind kritisch zu bewerten und bei einer Prüfung auf Wahrheitsgehalt im Vergleich mit der Realität nur schwer tragbar [9]. Aggregierte Modelle werden oft von agentenbasierten Modellen dominiert. Jedoch benötigen auch diese Modellart Annahmen und Vereinfachungen, deren Vorhersagekraft aufgrund der Komplexität von individuellem Handeln meist fragwürdig ist [11]. Serious Gaming bringt Menschen in die Evaluation ein und bedarf so keiner Verhaltensannahmen. Während analytische Ansätze meist von optimalem Handeln ausgehen, agieren Menschen eher heuristisch und pragmatisch. Zusammenfassend beobachtet Serious Gaming das Handeln des Menschen, nicht das Handeln der Modellannahmen.

Die verbreitetste Methode der Evaluation ist die Simulation, eine Nachahmung der Handlung eines realen Prozesses oder Systems über die Zeit [2]. Der Vorteil von Simulationen ist, dass sie eine aussagekräftige und kontrollierbare Umwelt besitzen, die Antworten zu verschiedensten Szenarien bieten kann und somit helfen kann, einen Prozess oder eine Aktion besser zu verstehen [2]. Jedoch ergibt sich auch hier bspw. die oben diskutierte Problematik der Verhaltensannahmen. Zudem benötigen Simulationen für die Berechnungen einen hohen Zeitaufwand und großes Fachwissen. Simulationstechnologien wie Hardware, Softwarelizenzen, Pflege der Technologie oder Ausbildungskosten sind relativ kostspielig [10]. Diese Einschränkungen führen in der Praxis oft zur bruchstückhaften Ausführung der Simulationen, um Zeit, Personal und Kosten zu sparen. Die Evaluation eines Mechanismus ist ein *Wicked Problem*. Dieses zeichnet sich durch besondere Charakteristiken aus: Es fehlen eine explizite Formulierung, Abbruchregeln oder eine eindeutige Überprüfung. Simulationen können bei diesen Charakteristiken problematisch sein [15]. Des Weiteren können Simulationen auch an fehlenden geeigneten Benchmarks leiden. Serious Gaming kann hier als eine Benchmark dienen und pragmatische menschliche Intelligenz darstellen. Diese menschlichen „Heuristiken“ können mit der Simulation verglichen werden und sie so ergänzen. Spiele benötigen keinerlei Wissen oder Annahmen über Verhalten oder die Dimension des Problems, welches von den Spielern im Spiel entdeckt werden können. Evolutionäre Strategien können sich ausbilden und die Evaluation profitiert von der Weisheit der Vielen, da Strategien, die von Fachleuten nicht in Betracht gezogen wurden, entdeckt werden können. Ein Beispiel: Zu Beginn des Zweiten Golfkrieges (1990–1991) wollte das US-Pentagon seine Reaktionsstrategien ausloten. Simulationen konnten keine ausreichend zufriedenstellenden Antworten liefern aufgrund von mangelnden Umweltinformationen. Man entschloss sich, ein kommerzielle Kriegsbrettspiel (Gulf Strike) einzusetzen, um spielerisch die meisten der später angewandten Strategien zu evaluieren und über diese zu verhandeln [3][20].

Feldstudien und Testläufe implementieren Artefakte in einem separierten Umfeld und finden so in einer realen Umwelt statt. Jedoch kann das separate Umfeld auch besondere Eigenschaften aufweisen, wie z.B. besondere Präferenzen auf einem regionalen Testmarkt, die das Ergebnis verzerren und die Allgemeingültigkeit in Frage stellen. Zudem sind Feldstudien sehr kostspielig, können negative Konsequenzen, wie negative Kundenerfahrung, haben und sind nicht reproduzier- und parametrisierbar. Laborexperimente hingegen implementieren Artefakte in einer Testumgebung mit wohldefinierten Umgebungseigenschaften und versuchen reale Situationen nachzubilden, wodurch sehr leicht verschiedene Alternativen getestet werden können. Jedoch sind Laborexperimente anfällig für verschiedene Verzerrungseffekte (vgl. im Folgenden [17]). Ein Experiment benötigt einen durchdachten Aufbau und Anreize für die Probanden, da ansonsten aus den Tests falsche Rückschlüsse gezogen werden können. Serious Games bieten einen zentralen Anreiz für die Teilnahme: Unterhaltung und Spaß. Die Probanden sind sich nicht dermaßen bewusst, an einer Studie teilzunehmen, wie sie es in einem Laborexperiment wären. Daher fühlen sie sich nicht dazu gedrängt, Antworten zu geben, die der Versuchsleiter vermutlich hören möchte. Dieses Problem ist als *Observer Expectancy Effect* oder *Versuchsleitereffekt* bekannt. Ein präsenter Versuchsleiter ist beim Serious Gaming nicht notwendig und eine große Zahl an Versuchspersonen kann, im Gegensatz zu Laborexperimenten, akquiriert werden. Des Weiteren haben Laborexperimente oft fast ausschließlich Studenten als Testpersonen und können so selten eine adäquate Grundgesamtheit nachbilden. Serious Gaming kann mit Quoten bestimmte Strukturen erreichen und dabei immer noch große Teilnehmerzahlen erreichen oder offen gestaltet sein. Es ist oft schwierig, Experten aus der Praxis für ein Laborexperiment zu

gewinnen. Serious Gaming kann dies erleichtern, weil es weder zeit- noch ortsgebunden ist und ad hoc online ausgeführt werden kann. Ein weiterer Nachteil von Laborexperimenten ist der sogenannte *Hawthorne Effect*: Probanden fühlen sich durch die Teilnahme an einem Experiment privilegiert und ändern ihr Verhalten – bspw. werden bei Leistungsmessungen auch in der Kontrollgruppe Anstiege gemessen. Bei Serious Gaming ist das Individuum sich nicht vollständig bewusst, an einem Experiment teilzunehmen, da es in erster Linie seinem Vergnügen an einem Spiel nachgeht. Des Weiteren könnte sogar der wissenschaftliche Zweck des Spieles nicht offenbart werden, sodass der Proband in vollem Maße im Glauben ist, lediglich ein Spiel zu spielen. Diese Vorgehensweise kann auch das sogenannte *Hypothesis Guessing* verhindern, wobei die Testperson versuchen könnte, die Hypothese zu raten und dementsprechend auf eine andere Weise darauf zu antworten. Allerdings hat eine von den Autoren durchgeführte Umfrage unter 70 Personen ergeben, dass Menschen eher gewillt sind, ein Spiel zu spielen, wenn sie wissen, dass es einem wissenschaftlichen Zweck dient. Zusammenfassend kann Serious Gaming die Projektion der realen Welt auf die Evaluationswelt verbessern, indem echte Agenten verwendet werden, die den Modellen in ihrer Verhaltensprognosekraft überlegen sind. Eine Kombination aus Serious Gaming, Experimenten und Simulationen ist denkbar und würde zu besseren Evaluationsergebnissen führen.

4 Leitfaden für eine Serious-Gaming-Evaluation

Das Ziel einer Evaluation ist es, die Güte eines Artefaktes zu bemessen, also die Qualität, Wirksamkeit, Effizienz, Anwendbarkeit und Nützlichkeit eines Artefaktes offen zu legen [5]. Dieser Prozess resultiert entweder in Vorschlägen zur Verbesserung oder autorisiert je nach Bewertung die tatsächliche Anwendung. Bild 1 zeigt einen Leitfaden für eine Serious-Gaming-Evaluation (vgl. [8]). Der Leitfaden orientiert sich an Pólyas [12] Ablaufempfehlungen zur Entwicklung von heuristischen Lösungen.

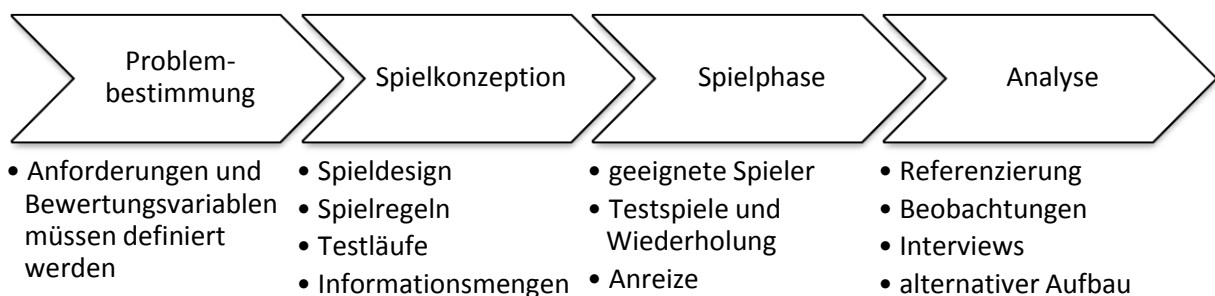


Bild 1: Leitfaden für den Ablauf einer Evaluation

Problembestimmung

Das Problem und der Zweck der Untersuchung müssen als Erstes bestimmt werden. Hierzu muss man Informationen über die Problematik einholen und sich mit der Materie auseinandersetzen. Die erarbeiteten Kriterien müssen dabei gründlichst untersucht und auf Relevanz geprüft werden [5]. Dies stellt die Vorarbeit für den folgenden Evaluationsprozess dar und ist sehr wichtig für dessen Erfolg, da auf der Basis der Problembestimmung das Serious Game und die Experimentumwelt konstruiert wird.

Spielkonzeption

Die Spielkonzeption besteht einerseits aus dem Spieldesign und andererseits aus den Spielregeln. Das Spieldesign umfasst die Schnittstelle zum menschlichen Spieler (Interface) sowie die Spielengine. Während das Interface bedeutend für das Spielvergnügen und die Benutzerfreundlichkeit ist, und somit entscheidend für die Akquirierung von Spielern, enthält die Spielengine die zugrunde liegende Umweltmodellierung. Die Spielregeln, wie z.B. Abspracheverbot, bestimmen die Interaktion zwischen den Spielern oder den Spielablauf. Es empfiehlt sich, Testläufe durchzuführen, um zu untersuchen, ob das Spiel den Anforderungen wie Benutzerfreundlichkeit gerecht wird. Testläufe sind ebenfalls eine Evaluationsmethode, die das neu geschaffene Artefakt, das Serious Game, bewertet, um eine möglichst hohe Aussagefähigkeit erzielen zu können.

Spielphase

Nach gründlichen Testläufen kann das Spiel veröffentlicht werden und die tatsächlichen Probanden das Spiel spielen. Bei der Auswahl der Spieler besteht eine Wahlmöglichkeit zwischen freiem Zugang und Quotenregelungen, um eine bestimmte Stichprobenstruktur zu erreichen. Ebenso muss die Notwendigkeit von Fachwissen für das Spiel bedacht werden und demnach müssen die Spieler ausgewählt werden. Den Spielern sollte die Möglichkeit gegeben werden, Proberunden zu spielen, oder das Spiel wiederholt spielen zu können. Dies kann nicht zielorientiertes Verhalten unterbinden und macht den Spieler mit dem Spiel vertraut. Anreize wie Preise oder Wettbewerb führen dazu, dass die Spieler sich aktiver und ambitionierter mit dem Spiel beschäftigen [1].

Analyse

In der Analyse wird das Ergebnis des Artefakts mit den Ergebnissen der Spieler verglichen, und andere Vergleichsmessungen wie ein theoretisches Optimum oder andere Artefakte, falls vorhanden, werden hinzugezogen. Des Weiteren können eine Analyse der Strategien der Spieler neue Informationen und Ideen bieten, die das Artefakt weiter verbessern können. Alternative Spelaufbauten können verwendet werden, um den Artefakt unter anderen Grundannahmen zu testen und die Gültigkeit der Evaluation weiter zu untermauern.

5 Distribution mittels Web 2.0

Web 2.0 als „Mitmachweb“ bietet Chancen für Serious-Gaming- Evaluationen; es können durch den Interaktionscharakter leicht sehr große Teilnehmerzahlen erreicht werden, was in klassischen Experimenten nur unter großem Aufwand und Kosten möglich wäre.

5.1 Soziale Netzwerke

Eines der Kernlemente des Web 2.0 sind Soziale Netzwerke wie facebook.com. Ein Vorteil dieser Netzwerke ist, dass persönliche Informationen über den Benutzer zur Auswertung verfügbar sind. Des Weiteren sind Anwendungen in Sozialen Netzwerken sehr erfolgreich, da durch Feeds, die den Kontakten der Person angezeigt werden, oftmals die Teilnahme an einer Anwendung oder deren erfolgreiches Absolvieren geteilt wird. Diese erfahren wiederum so von der Anwendung und können sie adaptieren. Diese Form der erleichterten Mund-zu-Mund-Propaganda führt zu einer schnellen Ausbreitung der Anwendung bei den Benutzern. Fogg [4] entwickelte hierzu die sogenannten Mass Interpersonal Persuasion (MIP) untergliedert in 6 Stufen

(siehe Bild 2). Folglich, sofern ein Serious Game unterhaltsam gestaltet ist, kann eine MIP induziert werden und eine große Probandenzahl erreicht werden. Um das Nutzerverhalten auf sozialen Plattformen näher zu untersuchen, wurden 70 Facebookmitglieder nach deren Gewohnheiten befragt: 79% der Befragten gaben an, dass sie Facebook-Anwendungen (Apps) verwenden, wobei 41 % diese oft bzw. sehr oft nutzen. Spiele waren dabei die am häufigsten regelmäßig benutzte Anwendungsgattung (23%), während 50% der Befragten generell, unabhängig von der Frequenz, Spiele auf Facebook nutzen. Die Veröffentlichungsfunktion, bei der z.B. Spielergebnisse auf der Startseite der Freunde als Feed angezeigt werden, wird von 71% verwendet. Diese Funktion ist ein Kernstück des MIP, da sie die soziale Distribution (Stufe 3) und einen schnellen Zyklus (Stufe 4) fördert. Der durchschnittliche Nutzer von Facebook hat 130 Freunde in seinem Netzwerk (Quelle: <http://facebook.com/press/info.php?statistics>, Stand: 08/11), die durch den Feed über das Spiel informiert werden. Somit verbreitet sich eine Anwendung durch das soziale Netz und wird ohne weiteren Eingriff der Entwickler beworben. Des Weiteren gaben 57% der Befragten an, dass sie schon einmal Freunde zu einer Anwendung eingeladen haben. Die soziale Verbindung zwischen den Spielern unterstützt nicht nur die Verbreitung des Spieles, sondern lässt auch den Einsatz im Spiel steigern. So bestätigten 38% der Befragten, dass sie schon einmal ein Spiel wiederholt spielten, um besser als Freunde abzuschneiden. Der soziale Wettbewerb kann einen tragenden Anreiz darstellen.



Bild 2: Stufen der Mass Interpersonal Persuasion [4]

5.2 Crowdsourcing

Crowdsourcing beschreibt eine Auslagerung von Arbeit an die Masse analog zum Outsourcing, bei dem Arbeit an Subunternehmer ausgelagert wird. Ein Beispiel ist Amazon Mechanical Turk (AMT), ein Micro-Task-Marktplatz, bei dem Aufgaben angeboten werden, die nur bedingt automatisiert vollzogen werden können wie bspw. die Identifizierung von Gegenständen auf Bildern oder das Schreiben einer Kurzbeschreibung [7]. Crowdsourcing kann auch in der Wissenschaft einen großen Beitrag leisten. So sind oft geringe Stichproben bzw. mangelnde große Zahlen problematisch. Durch Crowdsourcing können schnell und kostengünstig größere Teilnehmerzahlen für Studien bzw. Spieler für ein Serious Game erreicht werden. Die Kosten belaufen sich nach bei Amazon Mechanical Turk üblichen Tarifen auf ca. \$ 0,05 bis \$ 0,15 pro Spieler mit einer Spielzeit von drei bis zehn Minuten. Die Spieler werden zwar nur für einmaliges Spielen bezahlt, jedoch kann man davon ausgehen, dass bei Gefallen die Spieler das Spiel auch mehrmals spielen werden und somit auch mehrere Datensätze generieren. Des Weiteren besteht

auch die Möglichkeit, dass die via Crowdsourcing akquirierten Teilnehmer wiederum über Soziale Netzwerke das Spiel mit Freunden teilen und somit eine weitere soziale Kaskade auslösen. Zur Qualität haben Kittur et al. [7] gezeigt, dass bei geeignetem Studienaufbau die Güte der Ergebnisse durchaus zufriedenstellend für wissenschaftliche Ansprüche ist und viele Methoden durch die vergrößerte Grundgesamtheit übertrifft.

6 Fallstudie „Cloud Manager“

Um den Aufbau eines Serious Game und den Einsatz von Web 2.0 illustrativ darzustellen, wurde ein webbasiertes Managementspiel („*Cloud Manager 2*“) konzipiert und realisiert.

6.1 Szenario

Das Spiel bildet eine Entscheidungssituation unter Unsicherheit im Cloud Computing ab und bedient sich einem Forschungsszenario, das bei der Entwicklung von automatisierten Entscheidungsregeln bereits verwendet wurde [14]. Hierbei müssen Auftragsanfragen bearbeitet werden und die Aufträge umsatzmaximierend angenommen oder abgelehnt werden. Der Zeitpunkt der Auftragsanfrage ist in der Regel nicht identisch mit dem Startzeitpunkt des Auftrages. Es sind drei Grundressourcen (Speicher, Bandbreite, Rechenleistung) vorhanden, deren Kapazitäten nicht überschritten werden dürfen. Des Weiteren wurden Aspekte des Quality of Service (QoS) integriert, indem Kundentypen mit verschiedenen Prioritäten und unterschiedlichen Nebenbedingungen eingeführt wurden. Zudem hat der Spieler die Möglichkeit einen bereits zugesagten Auftrag gegen Zahlung einer Vertragsstrafe abzubuchen. Die Herausforderung des Spieles besteht also darin, dass Ressourcen verteilt werden müssen. Der Spieler muss also Ressourcen für spätere, eventuell lukrativere, Anfragen vorhalten, wobei der Spieler keine Kenntnis über zukünftige Anfrage hat und die ungenutzten Kapazitäten bereits verteilen kann (für weitere Informationen über das Szenario, siehe [13][14]).

6.2 Konzeption und Implementierung

Das Spiel *Cloud Manager 2* wurde gemäß dem Leitfaden in Abschnitt 4 entwickelt.

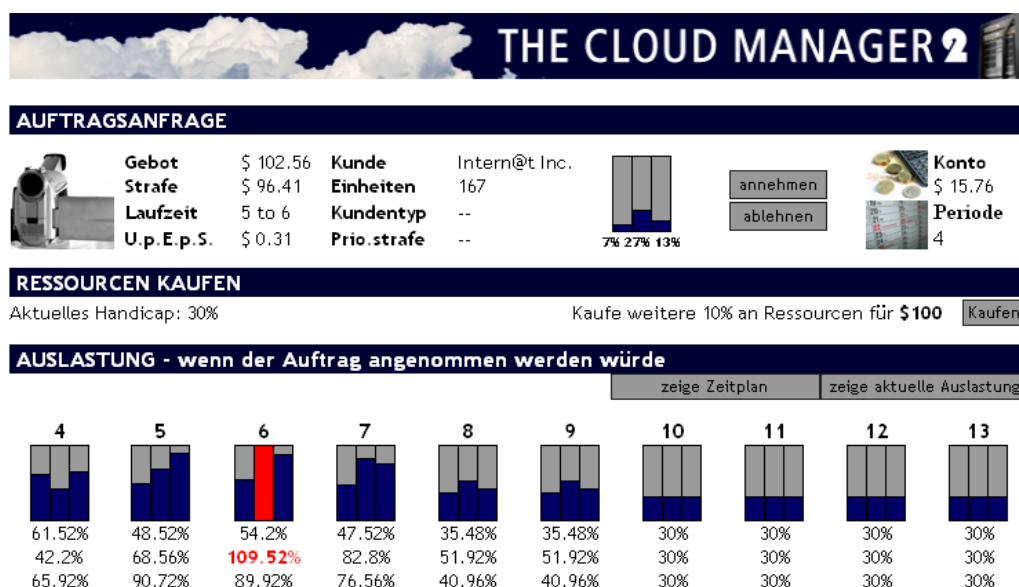


Bild 3: Screenshot des Spieldesigns

Problembestimmung

Aufbauend auf das Szenario stehen zwei Spielziele im Vordergrund: Erstens ein möglichst hoher Umsatz und zweitens eine hohe Annahmequote von Prioritätskunden. Dies sind Kennzahlen für sowohl Effizienz als auch Dienstgüte. Als Versuchsziel stehen, zum einen, die Ergebnisse der Spieler als Benchmark für den automatisierten Entscheidungsalgorithmus zur Verfügung und, zum anderen, können erfolgreiche Strategien der Spieler zur Verbesserung des Algorithmus verwendet werden.

Spielkonzeption

Das Spiel ist online spielbar und wurde mit HTML, php, JavaScript sowie AJAX in Verbindung mit einer MySQL-Datenbank programmiert. Zuerst wurde eine Version ohne die Nutzung von AJAX und JavaScript entwickelt (*Cloud Manager 1*), jedoch ergaben Rückmeldungen der Nutzer, dass eine verbesserte Dynamik des Spieles wünschenswert wäre und somit den Spielspaß und die Motivation der Probanden verbessern würde. Der Spieler muss über jeden Auftrag entscheiden, ob er ihn annehmen oder ablehnen will. Da das Spiel ohne Auslastung beginnt, wurde ein Handicap eingeführt, d.h. das Spiel beginnt mit einer Grundlast von 50% der Ressourcenkapazität. Dieses Handicap kann für einen geringen Geldbetrag abgebaut werden. Diese Funktion erhöht den Spielspaß und führt schneller zu Kapazitätsengpässen, die untersucht werden sollen. Ein durchschnittlicher Auftrag benötigt zwischen ca. 6% und 24% einer der Ressourcen. In realen Workloads ist dieser Wert wesentlich geringer, jedoch musste die Annahme getroffen werden, um das Szenario spielbar und interessant zu gestalten. Am Spielende werden alle angenommenen, aber noch nicht abgeschlossenen, Aufträge ausbezahlt, um den Spieleinsatz aufrecht zu halten.

Spielphase

Wie bereits erwähnt, wurde eine erste Version namens *Cloud Manager 1* entwickelt und durch interne Testläufe evaluiert. Diese Rückmeldungen wurden bei der Entwicklung von *Cloud Manager 2* miteinbezogen und diese Version wurde ebenfalls durch Testläufe evaluiert. Das Spiel wurde anschließend auf drei verschiedene Arten verteilt: Auf der sozialen Plattform facebook.com, durch Einladungen an Spieler mit Vorkenntnissen und durch den Crowdsourcing-Anbieter Amazon Mechanical Turk (AMT). Die Einladungen wurden in Newslettern und Foren für Cloud Computing verteilt, bei denen sowohl Wissenschaftler als auch Experten aus der Praxis gewonnen werden konnten. Das Spiel kann wiederholt gespielt werden – so besteht der Anreiz sich selbst zu verbessern und gerade in sozialen Plattformen besser als seine Kontakte abzuschneiden. Deshalb wurde in der Version für facebook.com auch eine Vergleichsfunktion mit Kontakten im Netzwerk eingefügt sowie eine Veröffentlichungsfunktion für Spielergebnisse. Als weiterer Anreiz wurde das Ergebnis des Spieler mit einer fiktiven Figur gleichgesetzt, die entweder negativ (z.B. „Managementfähigkeiten einer Kuh“) oder positiv (z.B. „Managementfähigkeiten eines Genies“) konnotiert waren.

Analyse

Die Ergebnisse der Spieler werden mit dem Optimum ins Verhältnis gesetzt, das man unter der Annahme vollständiger Informationen mit Hilfe einer gemischtganzzahligen Programmierung berechnen kann. Weitere Größen bzw. Verhältnisse sollen Informationen über die Annahmequote von Prioritätskunden oder die Abbruchkosten geben. Nicht nur die Ergebnisse des Durchschnittsspielers werden analysiert, sondern auch die Ergebnisse des besten Quartils und der

überdurchschnittlichen Spieler. In dieser Arbeit wird jedoch der Fokus auf Unterschiede der Distributionen, facebook und Amazon Mechanical Turk, gelegt.

6.3 Analyse der Implementierung in Web 2.0

Der größte Teil der 306 Datensätze wurde über Amazon Mechanical Turk (AMT) (49,3%) akquiriert, vor facebook.com (41,2%). Die restlichen 9,5% wurden über Einladungen und Foren erreicht. Hier zeigt sich die Eignung des Crowdsourcing um große Spielerzahlen zu erreichen, da über Einladungen nur 29 Datensätze gewonnen werden konnten, im Vergleich zu 277 über Web 2.0. Die 151 Datensätze auf AMT entstanden binnen 9 Tagen. 63,0% der Spiele wurden von männlichen Spielern gespielt (37,0% weiblich). Bei 7,2% der Datensätze war der Datensatz der erste Spielversuch. 60,8% hatten eine fachbezogene Ausbildung wie Wirtschaftswissenschaften, Informatik oder Mathematik und 81,4% waren Akademiker.

Um die Erfolgsfaktoren des Serious Game offenzulegen wurde eine Lineare Regression über die Spiele aus facebook und AMT durchgeführt (siehe Tabelle 1 links). Als abhängige Variable wurde das relative Ergebnis im Verhältnis zum theoretischen Optimum (=100) gewählt. Dadurch, dass bei der Einladungsgruppe keine eindeutige Nutzeridentifikation über eine eindeutige Identifikationsnummer wie bei facebook oder AMT möglich war, wurde diese Gruppe aus der Analyse entfernt. Die übriggebliebenen Gruppen facebook und AMT sind perfekt negativ korreliert, sodass eine AMT-Dummy-Variable zur Modellierung ausreichend ist.

	(1)	(2)	(3)		(4)	(5)
Konstante	74.591***	76.597***	75.472***	Konstante	54.040***	55.332***
AMTurk-DUM	-12.786***	-9.314***	-10.147***	5 für \$ 1-DUM	13.027***	12.073**
1. Spiel-DUM		-11.488***	-12.226***	1. Spiel-DUM		-1.515
Akademiker- DUM			2.585			
Fachbezug- DUM			5.522*			
Adjust. R²	10.31%	17.01%	17.98%	Adjust. R²	9.27%	8.75%

Abhängige Variable: Relative Ergebnis in % des theoretischen Optimums. Signifikanz: *** < 0,001 ** < 0,01 * < 0,05

Tabelle 1: Regressionsanalyse. Links: facebook und AMT (N=277). Rechts: AMT (N=151)

Regression (1) zeigt, dass AMT-Spieler signifikant schlechter abschneiden als Facebookspieler. In Regression (2) wurde eine weitere Dummy-Variable für den ersten Versuch einer Testperson eingeführt. Der Regressionskoeffizient von AMT verringert sich hierbei, was bedeuten kann, dass das schlechte Ergebnis teilweise damit zusammenhängt, dass Spiele auf AMT relativ öfter zum ersten Mal gespielt wurden (Korrelation von 0,32). Regression (3) zeigt, dass Nicht-Akademiker nicht signifikant anders als Akademiker abscheiden (p-Wert: 38,32%) und dass eine fachbezogene Ausbildung sich jedoch signifikant positiv auf die Performance auswirken (p-Wert: 2,26%). Spieler mit Fachbezug erreichen demnach ein um durchschnittlich 5,52 Prozentpunkte besseres Ergebnis.

Die AMT-Spiele wurden mit Hilfe von zwei unterschiedlichen Vergütungen, zum einen, € 20 für ein Spiel und, zum anderen, \$ 1 für fünf Spiele, angeworben. Pro Spiel wurde demnach der gleiche Betrag ausgezahlt. Für beide Vergütungsformen wurde ungefähr der gleiche Betrag aufgewendet. Bei der € 20-Vergütung wurde innerhalb von 7,5 Tagen 61 Spiele akquiriert, wohingegen bei der \$ 1-Vergütung binnen anderthalb Tagen 90 Spiele akquiriert werden konnten. Drei der insgesamt besten fünf Spieler stammen aus der Gruppe der \$ 1-Vergüteten. Regression (4)

zeigt eine Regressionsanalyse für die beiden AMT-Stichproben. Die \$ 1-Vergütung weist hierbei eine um 13,03 % bessere Performance auf. Dies könnte durch die vermehrte Wiederholung geschuldet sein. Daher wurde in Regression (5) ein Dummy für das erste Spiel eingeführt, der jedoch zu keiner signifikanten Veränderung des Ergebnisses führt. Der Vergleich zwischen den ersten Spielen der € 20-Gruppe und der \$ 1-Gruppe ergibt zudem, dass die \$ 1-Gruppe auch im ersten Spiel im Ergebnis überlegen ist. Eine mögliche Erklärung wäre, dass die Rüstkosten, also die Kosten der Einarbeitung in das Spiel, überproportional groß sind, sodass bei fünf Spielen ausreichend geringe Stückkosten, also Zeit pro absolviertem Spiel, realisiert werden können. Der ökonomische Gewinn pro Spiel, der aus der Differenz zwischen monetären Einnahmen und den Opportunitätskosten der Zeit pro Spiel besteht, würde somit steigen. Des Weiteren besteht bei Aufgaben, die eine längere Zeit benötigen, ein höheres Risiko versunkener Kosten, z.B. wenn ein Auftrag nach einem bestimmten Zeiteinsatz nicht zu Ende ausgeführt werden kann. Durch die längere Bearbeitungsdauer könnten Nutzer von AMT, die eher auf schnelle und qualitativ unsensible Aufgaben abzielen, um in möglichst kurzer Zeit möglich große Einnahmen zu realisieren, abgeschreckt sein. Im Gegenzug könnten Nutzer, die eher an gründlicher Arbeit an einem Projekt als an vielen kleineren Projekten interessiert sind, stärker angezogen werden. Vermutlich resultieren die unterschiedlichen Ergebnisse der beiden Gruppen aus der Selbstselektion, jedoch bedarf es weiterer Forschung, um diese Hypothese zu stützen.

7 Fazit

Serious Gaming ist eine alternative Evaluationsmethode, die Spiele als Umweltsimulation verwendet. Sie kann als Referenzpunkt dienen sowie eine Hilfe für die Verbesserung von Heuristiken durch Adaption von erfolgreichem Verhalten sein. Des Weiteren kann Serious Gaming menschliches Handeln erklären bzw. prognostizieren und somit Interaktion zwischen Mensch und Informationssystem analysieren. Die Methode benötigt keinerlei Annahmen über Verhalten wie in traditionellen Ansätzen, da reales Verhalten untersucht wird. Daher eignet sich Serious Gaming als gute Ergänzung zu traditionellen Methoden.

Eine große Grundgesamtheit gilt als Erfolgsfaktor für eine gründliche Evaluation. Wie in der Fallstudie gezeigt, können Soziale Netzwerke und Crowdsourcing hierbei sehr nützlich sein. Allerdings konnte auch gezeigt werden, dass die Ergebnisse sehr sensibel auf Änderungen des Versuchsaufbaus reagieren, da die Ergebnisse der Nutzer bei zwei verschiedenen Zahlungsschemen mit identischer Vergütung pro Spiel sich sehr unterschieden.

Die Aufgabe zukünftiger Forschung wird sein, die gespielten Strategien auf regelmäßige Handlungsmuster zu untersuchen, um erfolgreiche Strategien zu adaptieren. Das Spiel bietet ebenso die Möglichkeit, das Szenario zu ändern und die Reaktion der Spieler darauf zu messen. So könnte man auch die Reaktion auf Krisenszenarien (bspw. ein technisch bedingter überraschender Ausfall der Kapazitäten) untersuchen. Des Weiteren sollte sich zukünftige Forschung weiter auf das Potenzial von Crowdsourcing fokussieren. Hierbei sind einerseits die Untersuchung von Erfolgsfaktoren für den Versuchsaufbau und andererseits die Entwicklung eines Leitfadens für erfolgreiche Crowdsourcing-Implementierung von Interesse. Analog dazu sollten Erfolgsfaktoren für die Implementierung in Sozialen Netzwerken und Auslösung einer MIP untersucht werden.

8 Literaturverzeichnis

- [1] Babb, EM; Leslie, MA; Van Slyke, MD (1966): The Potential of Business-Gaming Methods in Research. *The Journal of Business* 39(4): 465-472.
- [2] Banks, J (1998): Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Wiley-IEEE, New York.
- [3] Dunnigan, JF (1992): The Complete Wargames Handbook. Morrow, New York.
- [4] Fogg, BJ (2008): Mass Interpersonal Persuasion: An Early View of a New Phenomenon. *Proceedings of the 3rd International Conference on Persuasive Technology*.
- [5] Hevner, A; March, S; Park, J; Ram, S (2004): Design Science in Information Systems Research. *MIS Quarterly* 28(1): 75-105.
- [6] Huizinga, J (1956): Homo Ludens: Vom Ursprung der Kultur im Spiel. Rowohlt, Reinbek.
- [7] Kittur, A; Chi, EH; Suh, B (2008): Crowdsourcing User Studies With Mechanical Turk. *Proceedings of the Twenty-Sixth Conference on Human factors in Computing Systems*.
- [8] Lang, F; Püschel, T; Neumann, D (2009): Serious Gaming for the Evaluation of Market Mechanisms. *Proceedings of the International Conference on Information Systems (ICIS 2009)*.
- [9] Lindenberg, S (1991): Die Methode der abnehmenden Abstraktion: Theoriegesteuerte Analyse und empirischer Gehalt. In: Esser, H; Troitzsch KG (Hrsg.), *Modellierung sozialer Prozesse*. Informationszentrum Sozialwissenschaften, Bonn.
- [10] McLean, C; Leong, S (2001): The Role of Simulation in Strategic Manufacturing. *Proceedings of the 2001 International Working Conference on Strategic Manufacturing*.
- [11] Neumann, D (2004): Market Engineering - A Structured Design Process for Electronic Markets. Universitätsverlag Karlsruhe, Karlsruhe.
- [12] Pólya, G (1945): How to Solve It. Princeton University Press, Princeton.
- [13] Püschel, T; Lang, F; Bodenstein, C; Neumann, D (2009): A Service Request Acceptance Model for Revenue Optimization: Evaluating Policies Using a Web Based Resource Management Game. *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences (HICSS-43)*.
- [14] Püschel, T; Neumann, D (2009): Management of Cloud Infrastructures: Policy-Based Revenue Optimization. *Proceedings of the International Conference on Information Systems (ICIS 2009)*.
- [15] Rittel, HW; Webber, MM (1973): Dilemmas in a General Theory of Planning. *Policy Sciences* 4: 155-169.
- [16] Rodriguez, H (2006): The Playful and the Serious: An Approximation to Huizinga's Homo Ludens. *Game Studies* 6(1).
- [17] Strohmets, DB (2008): Research Artifacts and the Social Psychology of Psychological Experiments. *Social and Personality Psychology Compass*, 2(2): 861-877.
- [18] Surowiecki, J (2004): Die Weisheit der Vielen - Warum Gruppen klüger sind als Einzelne. Goldmann, München.

- [19] Susi, T; Johannesson, M; Backlund, P (2007): Serious Games – An Overview. Technical Report HS- IKI -TR-07-001, Universität Skövde.
- [20] von Hilgers, P (2000): Eine Anleitung zur Anleitung. Das taktische Kriegsspiel 1812-1824. *Board Games Studies* 3.
- [21] Zyda, M (2005): From Visual Simulation to Virtual Reality to Games. *IEEE Computer* 38(9): 25-32.